Current View of Epidemiologic Study Designs for Occupational and Environmental Lung Diseases

Ira B. Tager

School of Public Health, University of California, Berkeley, Berkeley, California, USA

Epidemiologic studies long have played a role in the understanding of the effects of the general environment and various occupational exposures on the occurrence of acute and chronic diseases of the lung. This article is an overview of epidemiologic study designs that have particular relevance to studies of environmental and occupational lung disease. The application of times-series designs in the context of epidemiologic studies is discussed, as such designs have become widely used in studies of health effects ambient air pollution. The article emphasizes recent developments in the application of case–control study designs, many of which have had particular applications in epidemiologic studies related to environmental and occupational lung disease. These case–control designs offer efficient and valid alternatives for studies that in the past might have been conducted as more costly and time-consuming cohort studies. *Key words*: case–control studies, case–crossover studies, cohort studies, epidemiologic methods. — *Environ Health Perspect* 108(suppl 4):615–623 (2000).
http://ehpnet1.niehs.nih.gov/docs/2000/suppl-4/615-623tager/abstract.html

Epidemiologic methods have played an important role in the identification of environmental and occupational causes of lung diseases and respiratory morbidity/mortality in populations (*1–4*). Epidemiologic studies are concerned primarily with the distribution of causal determinants of disease in populations, but modern epidemiologic investigations increasingly have focused on more refined characterizations of subgroups of individuals out of which the cases actually arise (at-risk subsets) among exposed individuals within populations and on more refined characterizations of exposures. The expansion of the repertoire of objective markers of exposure, dose, and biologic susceptibility and response through advances in toxicology, molecular biology, and genetics has heightened this trend and is particularly relevant to occupational and environmental epidemiology that is related to lung disease. Epidemiologists also have made wider use of study designs using analytical strategies that have been largely applied in other disciplines (e.g., time-series analyses).

The goal of this review is to summarize the characteristics of *a*) selected epidemiologic study designs that have been developed/refined over recent years or *b*) more established methods, the application of which has increased as a consequence of a combination of convenience and sound theoretical underpinnings. The choice of the designs to discuss is motivated by the potential or proven usefulness of the designs in environmental and occupational epidemiology. A comprehensive review of methods for cohort studies recently has been published (*5*); therefore, this review focuses on cohort designs that either were not discussed or were discussed only briefly in that review. More emphasis is placed on case–control designs, as some of the newer designs are particularly relevant for occupational and environmental epidemiologic studies, and the potential efficiency of these designs can be of considerable advantage in terms of feasibility and costs. Moreover, improved understanding of the relation between case–control studies and failure time analysis in cohort studies (*6*) has provided a firmer basis for the validity of case–control design for causal inferences about disease occurrence in relation to particular exposures.

Cohort Study Designs

It is natural to begin with a discussion of cohort studies because in the realm of observational studies these designs most closely approximate the situation of an experiment— i.e., subjects free of disease are identified and occurrence of disease is observed among those exposed and not exposed to the substance under study. In contrast to a true experiment, the investigator has no control over who is exposed or other factors that might influence the outcome in the context of the particular exposure (i.e., confounders and effect modifiers). In a recent review of outcomes that can be studied with cohort designs, I divided these designs into two broad categories, life table-type and longitudinal (*7*). Life table cohort studies are characterized by their treatment of time and exposure in a manner tied closely to traditional life table methods of analysis. In general, exposure and person-time are summarized. Incidence (density) and cumulative incidence of a discrete disease outcome and their respective ratios are the principal outcomes of the life table-type cohort study. Inferences from theses types of cohort studies are restricted to population average effects. Longitudinal cohort studies, on the other hand, take explicit advantage of the potential for repeated measures of exposures and subject characteristics in cohort designs and make possible not only inferences on population average effects but also on individual heterogeneity, changes in processes over time, and repeated transitions back and forth between states of health and disease. In the review, I illustrated the usefulness of the longitudinal cohort study for five categories of inquiry (*7*). These are summarized in Table 1. Although the illustrative examples were drawn from a variety of disciplines, their applicability to studies of environmental and occupation respiratory disease is obvious. Two broad categories of cohort designs not considered in that review are quite relevant to epidemiologic studies of occupational and environmental lung diseases and will be the subject of the cohort studies section: time-series and panel studies, and multilevel cohort designs incorporating individual and between-group (ecologic) differences.

Contemporary epidemiologic studies of air pollution-related health effects rely on group-level (ecologic) assignments of exposure applied both to truly ecologic studies and to studies that use ecologic estimates of exposure and individual-level data. Therefore, it is useful to clarify what is and is not meant by an ecologic study [see Kunzli and Tager (*8*) and Morganstern (*9*) for more complete reviews] before I proceed to a discussion of specific study designs. In an ecologic study, all data are at the group level (e.g., daily incidence of death and some metric of air pollution for a given region on a given day); no individual level data are collected. Inferences about individuals from such studies are subject to what is called the ecologic fallacy (*10*). The ecologic fallacy is the bias that can occur when group data are used to make inferences about individuals and results from the mixing of between-group and within-group variability (*11*) and from group-level confounding and effect modification that do not have an immediate representation at the individual

Table 1. Types of analyses that can be conducted in longitudinal but not life table cohort studies.

Type of analysis	Example[a]
Estimation of effects at the individual level Estimation of individual heterogeneity Separation of cross-section versus longitudinal inference	Effects of respiratory illness on growth of lung function in children
Estimation of the effects of risk markers over time on disease outcome	Effect of hypertension on the occurrence of congestive heart failure
Estimation of the rate of change or change in level for outcomes that relate to disease natural history	Relationship between airways hyperresponsiveness and decline in lung function
Natural history of states of health with multiple occurrences or that oscillate between states of disability/disease	Probability of transition between being disabled and not being disabled in relation to age in an elderly population
Separation of cohort or period effects from the effects of age or calendar time	Effects of age and period effects on the decline in lung function in adults

[a]See Tager (7) for citations of specific examples.

level (12)]. Studies that obtain group-level data on exposure but individual-level data on important confounders and effect-modifiers are more properly considered individual-level studies, studies in which issues of exposure measurement error take on considerable importance (8).

Epidemiologic Time-Series Studies

Time-series studies have been one of the most frequently used longitudinal study designs over the past decade for the investigation of respiratory and other health effects related to ambient air pollutants because they are particularly useful for examining the effects of short-term fluctuations in air pollutant exposures on acute morbidity and mortality (13). In a general sense, a time series is a study based on a collection of observations made sequentially over time and applies to sequences of observations with intervals between observations that range from very short to very long, provided that a sufficient number of sequential observations are available in the series (14). Environmental epidemiologic time-series studies (ETSS) make repeated observations of exposures, outcomes, and relevant covariates over very short time intervals, usually days. This is in contrast to the more traditional repeated measures longitudinal epidemiologic studies (RMLES) (7) in which observations generally are separated by relatively long periods (e.g., years, decades). Although these latter studies are a time series in the strict sense of the term, the numbers of repeated observations in such studies are usually quite small. Thus, epidemiologic analyses of RMLES have tended to rely more on approaches related to failure time models (15,16). ETSS frequently use such models [e.g., Poisson regression (17)] but also use standard time-series analysis (18). One characteristic that distinguishes the ETSS design from the more traditional RMLES design is the need, in ETSS, to deal with potentially cyclical characteristics of time-dependent covariates as a confounding factor, i.e., the cyclical temporal character of a

particular factor induces the confounding (17). The fundamental comparison in ETSS is between fluctuations in outcomes (counts of events, deviations of measures of lung function from group or individual average levels) and fluctuation exposures (e.g., daily concentrations of air pollutants) after adjustment for confounding effects of time and weather. (An example of such confounding is the autumnal seasonal patterns of viral respiratory illness that coincide with seasonal increases in particulate air pollution, as might be observed in southern California.) The effect of the successful adjustment for confounding results in fluctuations of the outcome and exposure series around their respective mean values, which are stable, with stable variance [i.e., no time trend or periodic fluctuations in mean or variance, a so-called stationary time series (14)]. Figure 1 is an example from Samet and colleagues (19) to illustrate the successful removal of potentially confounding cyclical temporal trends in an outcome mortality series for a study of the effects of short-term fluctuations of ambient air pollutants on mortality.

In terms of the types of data collected, two general types of ETSS can be identified: eco-logic and individual-level (panel) studies. In the first type, data on exposure outcomes and covariates are obtained only at the group level. Examples are the many studies of daily cardiopulmonary and all-cause mortality or daily hospitalizations for cardiopulmonary morbidity and concentrations of ambient air pollution [summarized in Dockery and Pope (13)]. The outcomes are daily counts of events (usually for cities), the exposures are daily measures of ambient pollutant concentrations from central monitoring sites, and the covariates are time-related factors (e.g., day of week, season), meteorologic factors (e.g, temperature, relative humidity), and other potentially confounding factors that have obvious group-level interpretations (e.g., influenza epidemics). Sources of health-related data are secondary (e.g., death certificates, hospital admissions data) and are not based on any direct observations of individual subjects. These studies are true ecologic studies, the inferences from which are most properly limited to the population level. ETSS have particular appeal in the setting of studies of population-level health effects associated with daily fluctuation of ambient air pollutants because it can be assumed safely that, on a day-to-day basis, the populations under study do not change in terms of their distributions of important, unmeasured confounding and modifying factors (e.g., prevalence of cigarette smoking, patterns of health care that could affect patterns of hospital admissions, socioeconomic factors).

ETSS that make direct, daily measurements on individual subjects often are referred to as panel studies. Direct measures of outcome include such factors of lung function (18) and respiratory symptoms (20). Direct measures of potential confounders and effect modifiers include use of medications (20,21), cigarette smoking, and other non–air-polluting environmental exposures [e.g., fungal spores (22)]. Exposure to the relevant pollutants, on the other hand, most frequently are derived from central monitors and applied equally to all subjects, although
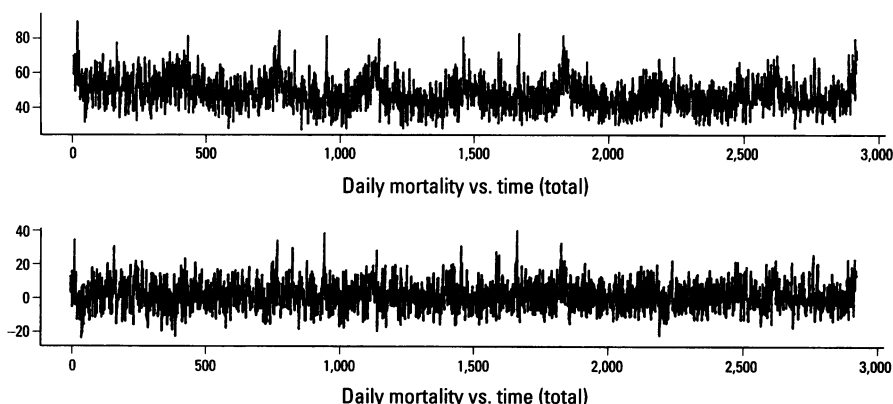

Figure 1. Top series is crude total mortality series for Philadelphia, PA, 1973–1980. Bottom series is the same mortality series after adjustment for long-term time trends in the data. Data from Samet et al. (19). Reproduced with permission of the Health Effects Institute.

panel studies allow for personal exposure measurements (23). Despite the availability of individual-level measurements, data from these studies often are grouped and the analyses are conducted as described for ecologic ETSS. In the case of outcomes such as lung function, panel studies permit the use of other modeling approaches that allow for inferences at the individual as well as the group level (16). Despite the usual focus on group-level inference and the use of group-level exposure data, panel studies are more properly considered individual-level studies in terms of issues of confounding and effect modification (8). The principal source of controversy in regard to such studies relates to the impact of exposure measurement error on the estimate of air pollution-related health effects (24,25). Panel studies have the same desirable property as ecologic ETSS of being able to assume that, on a day-to-day basis, individual-level unmeasured confounders and modifiers do not change.

Time patterns in the outcome, exposure, and covariate series can confound observed relationships between exposure and outcome (17) in ETSS and are cause for greater concern in these types of studies than in RMLES. In traditional RMLES, problems of time dependence generally are relatively simple— e.g., was a characteristic present or absent or at some particular level just prior to or in an interval prior to the observation of subjects? In ETSS, time patterns are more complex and can be fluctuations (seasonal patterns of deaths and air pollutant concentrations), day-of-week patterns of outcomes (decreased hospital admissions on weekends), or exposures (observations of higher weekend relative to weekday levels of ozone in the Los Angeles Basin), or long-term trends for ETSS that cover many years of daily data (long-term declines in air pollution concentrations and deaths from cardiovascular disease). Steps more specifically applied to ETSS are required in analysis to assure control of these time-dependent confounders (17).

Finally, it must be stated that despite their usefulness and importance for the study of short-term health effects related to air pollution, ETSS generally are not useful for inferences on the more long-term effects (26). One possible exception to this may emerge as methods are developed to address the problem of harvesting (27). This term arises in the context of observations of excess daily mortality in relation to daily increases in air pollution. Harvesting or mortality displacement refers to the possibility that only frail people, who are at high risk for imminent death, are affected by daily fluctuations in air pollution. Harvesting would be manifested by an increase in deaths immediately (or some short lag period of several days) after days with

increased air pollution and by decreases in deaths in the ensuing days due to the temporary depletion of the pool of frail, susceptible individuals. Concern was raised that the people who were actually dying were going to die very soon in any event and that the effect of increases in air pollution was only to hasten such deaths by a few days or weeks. If such a phenomenon is occurring, it would have far different implications from a public health viewpoint than an inference that the excess deaths were not imminent but were premature to some degree. Recent statistical methodological results suggests that inferences on this issue can be derived from ETSS (27).

Multilevel Designs

Cross-sectional comparisons between areas with different patterns of environmental pollution have been used to try to study the health effects of long-term (cumulative) effects of exposures increased levels of air pollution (28,29). Unfortunately, such studies suffer from such problems as temporality (does the measured exposure actually reflect the cumulative exposure that is inferred to have preceded disease onset and caused the disease), unmeasured or imperfectly specified confounders between populations, and, in the case of chronic disease, prevalence bias (i.e., those with the least lethal forms of the disease are most likely to be alive and be included in a cross-sectional study). To some extent true ecologic studies may have some advantages in this regard over cross-sectional studies based on individual-level data, as it is possible to construct long-term exposure records that are valid for a population (in contrast to specific individuals in the population) and to manage issues of confounding at the group level (e.g., group measures of socioeconomic status, race/ethnicity). Although such studies can provide valid inference at the population level (30), they are severely limited, at best, when inferences are required at the individual level.

Recently, an example of a multilevel design [the term multilevel, as used here, was proposed by Navidi et al. (25)] has been applied to the study of the long-term health effects of exposure to ambient air pollutants to take advantage of the cross-sectional contrasts that can be obtained in ecologic studies and the availability of individual-level data needed for inferences at the individual level (25,31,32). Only cross-sectional results have been published, but because the design of the study clearly is that of a classical RMLES (10-year cohort study), the design is presented in the section on cohort studies.

The basic design is fairly simple. Communities (n = 12) were selected to maximize contrasts between the various communities for various air pollutants and patterns of pollutants (25,31). Within each community,

a representative sample of children were selected for study. Health outcomes and other relevant individual-level data are collected on an annual basis. Annual air pollution data are obtained for each study community. Data on air pollution are available at the community level from central monitors and at the individual level based on microenvironmental modeling (25,33) and personal sampling.

The success of the design rests on identification of an analytical strategy that takes into account the within-community (i.e., the between-subject) and the between-community variability in outcomes and important confounders. A first-stage analysis model is fit [e.g., logistic regression (31), multiple linear regression (32)] for each community based on individual-level data for that community. This level of analysis provides a community-specific occurrence of disease or level of function that is adjusted for individual-level exposure as estimated from microenvironmental models (25). The second stage involves an ecologic model in which the outcomes are the community-adjusted estimates from the first stage analysis, and the exposures are the community-level exposures (e.g., community-specific averages of the individual-level exposure estimates) and, presumably, any other community-level covariates that can be estimated from averaging across individuals (e.g., prevalence of some disease or characteristic) (contextual variable) or are integral factors (e.g., weather conditions) that have no individual-level representation (34). Application of this design to the longitudinal data collected provides inferences on effects of long-term exposure on various outcomes that are not possible with single stage cross-sectional data.

Case–Control Study Designs

It may not be an overstatement to suggest that among nonepidemiologists, data derived from case–control studies often are treated with much skepticism, if not outright distrust. This skepticism undoubtedly is because in its classical formulation (35), the design seems counterintuitive. Cases and noncases (controls) are first identified and then their exposures are determined; this is quite the reverse of the more natural experimental situation in which an exposure is applied (experimental study) or experienced (observational cohort or longitudinal epidemiologic study) and the outcomes are observed subsequently. Since the 1970s, epidemiologists (36,37) and biostatisticians (6) have viewed the case–control study as a design that at its core is based on sampling from a cohort (the base, i.e., the real or theoretical cohort out of which the cases arise). In fact, most of the problems that relate to inferences derived from many case–control studies have more to do with problems of implementation

(e.g., selection bias of cases and/or controls, retrospective assessment of exposure, etc.) than with the theoretical validity of the various sampling strategies for different case–control designs. The attractiveness of the nested case–control design derives equally from the fact that the cohort from which the cases arise is known with certainty, selection biases can be minimized, and exposure assessment can be based on data available before the time the case arose. Therefore, to understand the designs to be discussed, a brief discussion of the case–base paradigm and risk set sampling is useful.

Miettinen (37–39) introduced the concept of the case base. He reasoned that case–control studies represented a special sampling from the person-time of some actual (population or primary base) or hypothetical population (secondary base). The term base refers specifically to the person-time experience out of which the cases arise. Ideally, under this so-called case–base paradigm, one obtains a complete census of the cases and a sample of the overall person-time of the cohort (controls). Validity in a case–control study, at least in terms of the selection of controls, therefore, depends on the proper identification of the base (population) out of which the cases arise and selection of an unbiased sample (with respect to the study base) of its person-time (controls) (40–42).

Parallel to the development of the case–base paradigm was the understanding gained by biostatisticians of the relationship between case–control sampling and failure times in a survival analysis (6) and the formulation of the incidence density case–control study (36). In this context, at the time of occurrence of a new (incident) case of disease, controls are sampled at random from the appropriate large population (base), which leads to the same conditional analysis associated with matched case–control studies (6). An extension of this approach led to control sampling strategies from defined cohorts (in contrast to an open, dynamic population). In one strategy, risk sets are sampled at the time of occurrence of the cases; this strategy forms the basis for the nested case–control study. In the case–cohort design, a subcohort is selected at random from the entire cohort at the start of the follow-up of the cohort and cases are identified from the entire cohort as the cohort moves through time (43).

Risk Set Sampling and the Nested Case–Control Design

The concept of a risk set is illustrated in Figure 2. Risk set sampling designs are related to methods used in the familiar Cox proportional hazards model for full cohort data (44). In this latter method, at each time a new case appears a risk set is formed comprised of the case and all other cohort members who are still at risk at the time the case appeared (i.e.,

the controls). A sample from the risk set would include the case and a sample (1 or more) of the eligible controls in the risk set at the time the case appears. Each time a case appears, a new risk set is formed and sampled. When risk sets are sampled in this manner, cases and controls are matched on time and also can be matched on important other confounders (e.g., age, sex, etc.). The final data set produced by this sampling strategy has the structure of a matched case–control study and the analysis follows a similar form (44). Langholtz and Goldstein (44) present a simplified explanation of the connection between this sampling strategy and the full cohort situation. A more in-depth summary is offered by Borgan and Langholtz (45). Estimation of the cumulative baseline incidence (hazard) of disease in the full cohort also is possible with these designs. Finally, in the usual nested case–control design cases and controls receive the same sampling weights but other risk set sampling designs can be implemented by specification of the appropriate sampling weights (44) (see below under "Staged Sample Case–Control Studies").

In a nested case–control design, cases are identified from a cohort that currently is under observation or from a cohort that is constructed retrospectively. At the time of occurrence of each case (e.g., age, time from start of employment, time from inception of cohort), one or more controls are selected from cohort members still at risk for the outcome at the time the case is identified. These controls also can be matched to the case on known important confounders (e.g., sex, smoking history, etc.). Such designs are particularly efficient when the outcome of interest is rare (e.g., leukemia in an occupation cohort) or the cost of the collection and processing of data is too expensive for implementation in an entire cohort (e.g., analysis of blood specimens for a biomarker, coding or summary of complex job matrices to assess occupational exposures, etc.). The odds ratios estimated from such studies are estimates of the average incidence density ratio for the cohort with no rare disease assumptions required (46). Moreover, the matching on time removes any assumptions about the prevalence of exposure during the life of the cohort (46).

A Norwegian study of the effect of exposure to NO_2 on the occurrence of bronchial obstruction in children less than 2 years of age offers a relevant example of a nested case–control study that illustrates the advantages of the design (47). A birth cohort of 3,754 subjects was assembled over a 1-year period and subjects were evaluated at 6-month intervals over the first 24 months of life. Children who met the criteria for bronchial obstruction and could be contacted (overall prevalence = 6.8%; 84% of all cases

could be contacted) were matched with children free of bronchial obstruction who were born next in time relative to each case. Indoor and outdoor NO_2 measurements with Palmes tubes in the home were made for a 2-week period only for the nested case–control subjects. Conditional logistic regression was conducted controlling for sex of child, birth weight, parental asthma, length of breast feeding, etc. Cases were relatively uncommon in the cohort (6.8%), and logistically expensive exposure assessments needed to carried out for only a small fraction of the cohort.

Staged Sample Case–Control Studies

The efficiency (precision and/or cost) of case–control studies at times can be improved by using more complex sampling designs than the simple nested case–control study (6,48). These designs are useful in situations in which a well-defined cohort either does not exist or has not yet been studied. Following are two types of examples.

Two-stage case–control studies. Two-stage case–control studies are one type of the more general two-stage designs (49). The general design is useful for those circumstances in which exposure and outcome data may be available for a large group of individuals, although information on important modifiers and confounders is not available and it is not feasible (cost, logistics) to obtain the missing data on all subjects. This design is illustrated in Table 2 (6). In this example, a hypothetical case–control study of 1,000 subjects was conducted (Stage I) to investigate the relationship between lung disease and employment in a particular factory. Employment in a factory is rare in this particular community. Cases and controls were obtained at random from the community. Employment status was known from existing data, but information on smoking was not available. Funds were not
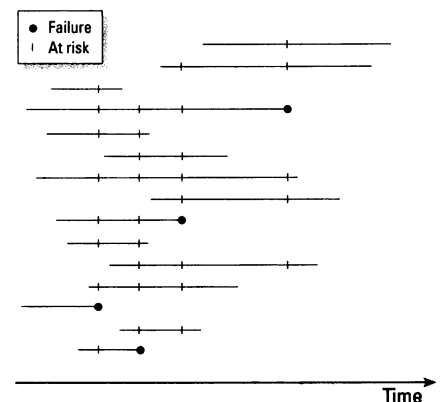

Figure 2. Hypothetical cohort. Each line represents a subject's time on study. ı represents an at-risk subject at the time a case appears; • represents the occurrence of a case. Data from Langholtz and Goldstein (44). Reproduced with permission of *Statistical Science.*

sufficient to interview all subjects. As exposure (factory) was uncommon, all exposed cases and controls were interviewed, but only a sample of unexposed cases and controls were interviewed (Stage II). An analysis restricted to only the Stage II sample would be less precise because of the smaller number of subjects and also would be potentially biased.

The essential feature of the analysis of such a design is to use the information on the disease–exposure relationship contained in the full study group (the 1,000 subjects in the above example) and the confounder information available for the Stage II sample. The various analytical strategies have two common characteristics: a confounder-adjusted estimate of the exposure–outcome association that is based on data from the second-stage sample; and adjustment of the exposure–outcome estimate and its variance for the method of sampling used to obtain the second-stage sample. Methods also have been developed to estimate the optimal second-stage sampling strategy in situations for which the goal is to optimize precision, or cost, or the situation in which the sizes of the first- and second-stage samples are fixed (50). This design strategy also can be applied when

subsample validations are conducted in the context of a case–control study (6). In this application, the assumption is made that the imperfectly measured variable (e.g., job matrix) is a "perfect" surrogate for the "gold standard" measurement—i.e., conditional on the true measurement, the surrogate has no association with the disease outcome.

Counter-matched studies. The simple nested case–control study uses only information for the cases and controls actually used the case–control study; none of the other cohort information is used. One consequence of this strategy is a lowering of the precision of the exposure effect estimates by using this subsample rather than the full cohort. Some of this loss in efficiency also is because matching on potential confounders in nested case–control studies (in fact, in all matched case–control studies) also leads to some level of matching on exposure. Such exposure-concordant pairs (risk sets) do not contribute to the estimation of the exposure effect. The countermatching design, which derives from risk-set sampling (44), seeks to remedy these shortcomings by stratified sampling of exposures (not confounders, as usual with matching) and by adjusting a subject's relative risk

for the sampling strategy from the strata of exposure (in the simplest situation, exposed and unexposed) in each risk set (44,51)—i.e., the subjects are weighted relative the probability of being sampled from a given exposure stratum in the risk set at the time the case appears. The term countermatching comes from the design with one control for each case and only two exposure strata (e.g., exposed and unexposed). In this case, the control is selected from the exposure stratum opposite from (counter to) that of the case. When more than one control per case is to be selected, subjects are selected such that the number of subjects per exposure stratum is equal (say, m_l). Therefore, each stratum will contain m_l controls except the case stratum that will contain $m_l - 1$ control plus the case (45). The analysis form is a weighted conditional logistic regression (45). Langholz and Goldstein (44) and Langholz and Clayton (51) provide a number of situations in which such a design might be useful and provide examples. These are summarized in Table 3.

Single-Case (Case-Only) Designs

This class of designs is based on the general concept that cases can serve as their own controls. All of these designs share a common need to use some prior theory or set of assumptions about exposure distribution to replace information usually supplied by an independent control group (52). (In the unusual case–control study, the control exposure distribution is used to estimate the underlying exposure distribution.) Thus, the ultimate validity of such designs depends on the validity of the assumptions about the exposure distribution (52,53). These designs have application in genetic analysis (54), but focus here is on designs and assumptions relevant to etiologic epidemiologic studies of environment and occupational exposures.

Table 2. Data from hypothetical two-stage case–control study.[a]

Exposure	Stage I	
	Cases	Controls
Factory workers	40	20
Employed elsewhere	460	480
Total	500	500

	Stage II			
	Cases		Controls	
Exposure	Factory	Elsewhere	Factory	Elsewhere
Smokers	30	37	10	42
Nonsmokers	10	133	10	128
Total	40	170	20	170

[a]Data from Breslow (65).

Table 3. Situations in which countermatching designs may be helpful.[a]

Situation	Goal	Comments
Crude exposure surrogate available on all subjects; more detailed exposure available on limited sample	Assess the effect of the detailed exposure variable	More efficient (↓ variance) than nested case–control The more accurate the surrogate the closer the efficiency to the full cohort relative to a simple nested design Exposure strata may be optimal when based on case distribution of true exposure and equal-size strata
Exposure data available on all subjects and confounders to be collected on a subset	Assess confounder-adjusted exposure effect	Confounder effect and variance seemed to be captured better with 1:3 countermatched design than with 1:1 Confounder effect may not be estimated any more accurately or precisely than simple nested design, but the main exposure effect more is accurate and the variance smaller
Exposure data available on everyone, and another covariate is collected on the sample	Investigate interaction between exposure and covariate	For most situations in which exposure is associated with disease, countermatching may be more efficient for the estimation of the interaction parameter (↓ variance)
Exposure 1 data available on all subjects and exposure 2 data available only on a sample of subjects	Investigate exposure 1 and main effect of exposure 2	Low efficiency Proposes strategy for use of additional randomly sampled controls, which improves the efficiency of estimation of the exposure 2 while retaining the countermatching efficiency for the primary exposure (exposure 1)

[a]Data modified from Langholz and Goldstein (44).

Case–crossover design. This design was introduced by Maclure in 1991 (*53*) "to assess the change in risk of a rare acute event during a brief 'hazard period' [*sic*] following transient exposure to a determinant of event onset" (*55*). Reasoning from the case–base paradigm (*37*), Maclure viewed the case–crossover design as the counterpart to a cohort study in which subjects crossed over between periods of exposure and nonexposure. He viewed the design as analogous to crossover designs used in treatment studies except, as is true for all observational studies, crossover times between risk and nonrisk periods are not random (*53*). The obvious advantage of such a strategy is the complete control for time-invariant potential confounders or the near-complete control for confounders that change slowly over time (*52,53*).

In keeping with the case–base paradigm and risk set sampling, a complete census of the risk period immediately before the occurrence of the event under study [e.g., heavy

Figure 3. Examples of three different approaches for the choice of controls in a unidirectional case–crossover design. (*A*) Single control time intervals. (*B*) Multiple control time intervals. (*C*) Usual frequency over some predefined interval prior to the case interval. Data from Mittleman et al. (*55*). Reproduced with permission of Oxford Press.

physical exertion in the hour before a myocardial infarction (MI) in Maclure's original paper (*53*) or the 48-hr period prior to death in studies of mortality associated with air pollution (*56*)] and an appropriate sample of comparable time periods prior to the event are obtained for each subject. The sampling strategy is illustrated in Figure 3. The choice of control time periods is governed by the availability and representativeness of the exposure data for the control period (*55*) and a stable (over time) exposure distribution. In the original design proposed by Maclure (*53*), exposures can be either point exposures or exposures of short duration. Each such exposure has an induction period (shortest time between exposure and onset of outcome) and a duration of effect (duration of effect after the end of induction period). Some direct knowledge or assumptions about both the induction and effective periods are critical to prevent carryover effects from control periods into the time of the risk period immediately preceding the event. Each subject then is treated as a matched pair composed of an exposure period immediately before the event and an exposure period not associated with an event. Thus, the data analysis follows that for traditional matched case–control studies, i.e., Mantel-Haenszel estimates or maximum likelihood methods that include conditional logistic regression (*52,57*). As the control data (exposure times) are measured in person-time units, this design estimates the average incidence rate (density) ratio (*53*). Willich and colleagues (*58*) compared the results of a conventional case–control study and those of a case–crossover study to address the question of whether the occurrence of an MI is related to recent physical activity (*58*). Only 270 of 882 cases experiencing MI could provide adequate information on physical activity for the case–crossover design. Nonetheless, the odds ratios for the occurrence of physical activity in the 1-hr prior to MI were identical for the two designs (2.1; 95% confidence interval, 1.1–3.6).

A number of potential sources of bias are possible for the case–crossover design (Table 4).

One that is unique to the original design ("unidirectional") is the assumption that the exposure distribution and subject-specific confounders are stable over time. A corollary is that subject-specific confounder–exposure relationships also do not change over time. Carryover effects from the control time into the time period immediately preceding the event also lead to biased estimates of exposure effect. Information bias can result from the need to obtain information about the control exposure in a manner that is different from that used to obtain the exposure information just prior to the case [e.g., the decision to use usual exposure to define the control time compared to using exposure in some finite time period prior to the event (*55*)].

In addition to the usual sources of selection biases in case–control studies (e.g., subject willingness to participate related to exposure or outcome), the failure of the assumption of no time trends in exposure leads to a form of selection bias. If there is a temporal trend in exposure, control time systematically may have greater or lesser exposure than case time, with the direction dependent on the nature of the temporal trend (*25,59*). Navidi and colleagues (*25*) developed the bidirectional case–crossover design to address this problem in studies of environmental exposures. Assessing environmental exposures is sometimes easier than assessing behavioral exposures (e.g., exercise in Maclure's example) for two reasons: reasonably accurate information about past levels of exposure often is available, and levels of exposure are not affected by the outcome in the subject. Navidi et al. called the design bidirectional because the control exposures could be sampled from times before or after the event of interest. In a simulation study based on temporal patterns of ambient particulate mass < 10 μM in diameter (PM_{10}), the authors demonstrated that unbiased estimates of effect could be obtained when biased estimates were obtained for a unidirectional case–crossover design (Table 5) (*25*). Moreover, the variances were small for the bidirectional design. One criticism raised
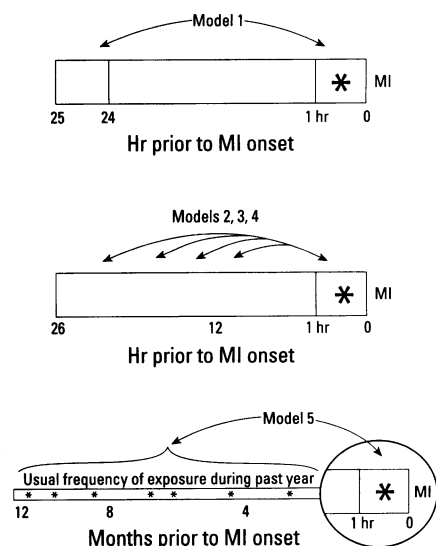
Table 4. Case-only studies.

Type of study	Features of sampling of control exposure	Rationale	Potential threats to validity (sources of bias)
Case–crossover			
Unidirectional	Selected from relevant time period(s) prior to case exposure period	Controls for subject confounders that remain fixed over time (e.g., sex, ethnicity)	General and confounder-specific time trends in exposure Applies to both uni- and bidirectional sampling
Bidirectional	Selected from relevant time periods before and after case exposure period	Potential to control for temporal trends in exposure	Time trends in confounder–exposure risks Carryover effects of exposure Case selection bias Information bias
Case–specular (developed specifically to study effects of magnetic fields)	Defined by wire code of hypothetical residences located in a virtual situation in which either the position of the power line is switched around the center of the street of the case residence	Avoids bias in control selection in relation to important neighborhood characteristics	Failure of the assumption of symmetrical probably of a given wire code Bias in wire coding

about this method (*25*) was its failure to account for carryover effects in the subject (e.g., a person who experienced an MI a number of days after an air pollution exposure is not the same person he/she was in before exposure in terms of underlying susceptibility to MI, as might be evidenced by decreased heart rate variability). Bateson and Schwartz (*60*) noted that the simulations carried out by Navidi et al. did not address the issue of confounding by the omission of variables that are the source of seasonal variations in morbidity and mortality associated with short-term fluctuations in ambient air pollutants. The simulation undertaken by these investigators demonstrated that not all bidirectional control sampling strategies led to unbiased estimates over the eight different patterns of seasonal and time-trend scenarios used by the authors and that estimates could be unbiased for some types of season/time-trend patterns and biased for others. Moreover, they observed that there were substantial losses in efficiency (increased variance) relative to standard Poisson regression approaches even for those bidirectional crossover designs that gave unbiased results. The work of Bateson and Schwartz (*60*) indicates that the validity of results from a bidirectional design clearly cannot be assumed, in the absence of data from other design strategies, when complex temporal exposure patterns exist (such as those encountered in many studies of ambient air pollution). Moreover, their simulation indicates that appropriate control sampling strategies could differ between study areas with differing patterns of temporal confounding.

Neas et al. (*56*) used the case–crossover design to reevaluate the relationship between daily mortality and daily fluctuations in total suspended particulates (TSP) over the years 1973–1980. They evaluated uni- and bidirectional control strategies (7, 14, 21 days). The period of exposure risk was the 48 hr prior to death (based on previously published data). Table 6 demonstrates two important points: choice of the control interval affects the risk estimate (multiple day control periods produced estimates closest to that obtained in Poisson regression [relative risk = 1.069] based on full time series), and the bidirectional

design was able to control for the effects of seasonal fluctuations in TSP in contrast to the unidirectional design.

Case–specular designs. This design was developed by Zaffanella and colleagues (*61*) specifically to address the problem of control selection (e.g., selection bias due to non-response associated with important neighborhood characteristics) in studies of health effects related to exposure to electric and magnetic fields. Instead of selecting an actual control and wire coding for the case residence, the case–specular method compares the wire code of hypothetical residences (specular residences—specular means mirror or reflection) located in a virtual position in which either the position of the residence or the power line is switched around the center of the street (Figure 4). The specular residence matches the neighborhood characteristics of the case house, but it may have a different wire code. Most important for this design, if the association between wire codes and a health effect resulted from wire code acting as a proxy for some neighborhood/street factor (e.g., air pollution, socioeconomic status), then the distribution of wire codes should not differ between case and specular residences. This latter point is the critical assumption for the method (*52*) and is tested through the explicit assumption that the probability of encountering a residence with an actual wire code X and a specular wire code Y is the same as the probability of encountering a residence with actual wire code Y and specular code X (*61*). This assumption of symmetry can be assessed only through a neighborhood survey or the use of actual controls. Two other assumptions of this approach are that residences on the same side of the street as power lines are not systematically different (i.e., on other risk factors) from residences on the opposite side of the street, and that coding of case residences and their speculars is done in an unbiased manner. A test of the first assumption was made for one study site and the assumption was supported; however, the authors acknowledge the need for tests in various neighborhood configurations to determine the ease with which this critical assumption can be met. To avoid bias in the coding of wire codes, the use of a control household is desirable (to blind coders); in a pilot study without

blinding and explicit coding runs this did not present a problem. Clearly more testing of with this design and the likelihood of meeting its critical assumptions is required before it can be recommended as a sole source of data for studies of health effects of electrical and electromagnetic fields.

General comments about case-only designs. The case–specular design has been developed for the specific problem of measuring environmental exposure. Therefore, general comments refer largely to the case–crossover design. Although the case–crossover design can be recommended for use in environmental and occupational respiratory studies in those situations that conform to the original temporal assumptions defined by Maclure (*53*), it must be acknowledged that the optimal strategy (in terms of lack of bias and maximum precision) for sampling control time is difficult to identify, even for a given exposure–outcome scenario (*55*) and extensions to exposure–outcome scenarios outside a given study would

Table 6. Comparison of bidirectional and unidirectional case–crossover designs. Adjusted odds ratios for the association of daily mortality with a 100 μm/m³ increment 48-hr average total of suspended particles, Philadelphia, PA, 1973–1980.

Control periods[a]	Without seasonal adjustment[b]	With seasonal adjustment[c]
Before and after case period		
7 days	1.105	1.104
7 and 14 days	1.050	1.048
7, 14, and 21 days	1.056	1.055
After case period only		
7 days	1.123	1.106
7 and 14 days	1.073	1.047
7, 14, and 21 days	1.072	1.041

From Neas et al. (*56*); estimate from Poisson regression based on full times series = 1.069 (95% confidence interval, 1.043–1.096). [a]Case period for exposure = 48-hr average TSP ending on the day of death. [b]Also adjusted for 24-hr average temperature, dew point, winter temperature, day indicator. [c]Adjustment for season with four sine and cosine functions and with periods of 365 and 120 days.
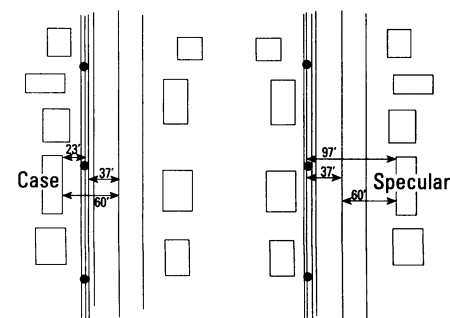

Figure 4. Diagram of case and specular residences in relation to power lines on a street. The specular residence is the mirror image of the case residence with respect to the center of the street. Distances to power line for the two residences are shown. Data from Zaffanella (*61*). Reproduced with permission of Lippincott Williams & Wilkins.

Table 5. Simulation study to compare standard case–crossover and bidirectional case–crossover estimators.[a]

Estimator	True value	Mean[b]	SD	Nominal SE
14 days prior for control day	0.1000	0.0625	0.1707	0.1730
28 days prior for control day	0.1000	−0.0220	0.1708	0.1701
Random day before and after case day	0.1000	0.0993	0.1330	0.1412
All days (n = 363) prior to failure	0.1000	0.4894	0.0710	0.0856
All days before (n = 363) and after (n = 363) failure	0.1000	0.0995	0.0934	0.0988

SD, standard deviation.
[a]Data from Navidi et al. (*25*). [b]Mean of 1,000 replications. For each estimator, this represents the square root of the average of 1,000 replications of asymptotic variance.

be hazardous. Therefore, any study that uses the case–crossover design should include several formulations of control exposure time to test the robustness and precision of the results. The work of Bateson and Schwartz (*60*) and a recent analysis from Korea (*62*) clearly indicate that temporal trends in exposure still pose a serious challenge to the validity of effect estimates derived using the case–crossover design. Thus, using the case–crossover design in situations in which it is known or suspected there are trends in exposures over time is best reserved for the corroboration of analyses that can more directly account for the effects of time trends. The case–crossover design does, however, have an advantage over the ecologic time series studies against which it has been evaluated because it permits evaluation of individual-level covariates, an option that does not exist for the truly ecologic time series studies. This option permits exploration of issues such as susceptible subgroups, which cannot be addressed easily in ecologic time series studies. In addition, if some of the problems of temporal confounding can be overcome, the design is more efficient than a panel study because expensive follow-up of panels of individuals may not be necessary.

Greenland (*52*) has provided a comprehensive discussion of issues that relate broadly to case-only designs. Several of his points are worth noting here. In general, the matching implicit in case–crossover designs can lead to decreased precision of estimates, since the close matching of these designs tends to produce exposure-concordant case-"pseudo-control" (Greenland's term) sets which are not used in the match analyses. Any misclassification of exposure is exaggerated in situations in which exposures are highly correlated (*63*), as they will tend to be in the case–crossover design. Moreover, misclassification could be differential within each case-pseudo-control set if different methods (metrics) of exposure are used for the case risk period and the pseudocontrol" period [see Mittelman (*55*) for examples] or because past exposures are less well-documented or recalled than current exposures. Finally, Greenland raises a subtle and important point that makes it difficult to compare results from case–control and case-only designs. He notes that exposure coefficients from case–control studies represent covariate-specific log-odds ratios that have the interpretation of average risk divided by average survival probability, and those from a case-crossover design have the interpretation of average subject-specific log-odds ratios (*52*). These two types of log–odds will have the same interpretion only if it can be assumed that all individual log–odds ratios are equal to a constant (equal susceptibilty) (*64*), a very strong assumption in most cases.

A Final Comment about Case–Control Studies

Modern understanding of the case–control study and its relation to the cohort study has led to a broad expansion of situations in which one or another variant of case–control studies can provide efficient and valid designs for investigation of the effects of environmental and occupational exposures on respiratory health. Some of these designs have applications in particular situations (e.g., case–specular), but in other situations, the strengths and pitfalls (e.g., case–crossover, countermatching) of the design have yet to be fully understood before the full range of their potential application is known. Ongoing methodologic research should provide this guidance in the not too distant future. Nonetheless, it should be clear that case–control designs deserve strong consideration as design options to address research questions that in the past might quickly have led to more costly and time-consuming cohort studies.

REFERENCES AND NOTES

1. Dockery DW, Speizer FE, Stram DO, Ware JH, Spengler JD, Ferris BG. Effects of inhalable particles on respiratory health effects of children. Am Rev Respir Dis 139:587–594 (1989).
2. Abbey DE, Petersen FF, Mills PK, Kittle L. Chronic respiratory disease associated with long term ambient concentrations of sulfates and other air pollutants. J Expos Anal Environ Epidemiol 3:99–115 (1993).
3. Doll R. Mortality from lung cancer in asbestos workers. Br Med J 12:81–86 (1955).
4. Siemiatycki J. Exposure assessment in community-based studies of occupational cancer. Occup Hyg 3:41–58 (1996).
5. Samet JM, Munoz A (eds). Cohort Studies. Epidemiol Rev 20:1–136 (1998).
6. Breslow N. Statistics in epidemiology: the case-control study. J Am Stat Assoc 91:14–28 (1996).
7. Tager IB. Outcomes in cohort studies. Epidemiol Rev 20:15–28 (1998).
8. Künzli N, Tager IB. The semi-ecologic study in air pollution epidemiology—a valid design as compared to ecologic studies. Environ Health Perspect 105:1078–1083 (1997).
9. Morganstern H. Ecologic studies. In: Modern Epidemiology, 2nd ed (Rothman KJ, Greenland S, eds). Philadelphia:Lippincott-Raven, 1998;459–480.
10. Morgenstern H. Uses of ecologic analysis in epidemiologic research. Am J Public Health 72:1336–1344 (1982).
11. Piantadosi S, Byar D, Green S. The ecological fallacy. Am J Epidemiol 127:893–904 (1988).
12. Greenland S, Morgenstern H. Ecological bias, confounding, and effect modification. Int J Epidemiol 18:269–274 (1989).
13. Dockery DW, Pope CA. Acute respiratory effects of particulate air pollution. Annu Rev Public Health 15:107–132 (1994).
14. Chatfield C. The Analysis of Time Series—An Introduction. New York:Chapman and Hall, 1984;1–286.
15. Thomas D. New techniques for the analysis of cohort studies. Epidemiol Rev 20:122–134 (1998).
16. Diggle P, Liang K-Y, Zeger S. Analysis of Longitudinal Data. New York:Oxford University Press, 1994;253 pp.
17. Schwartz J, Spix C, Touloumi G, Bacharova L, Barumamdzadeh T, le Tertre A, Piekarksi T, Ponce de Leon A, Ponka A, Rossi G, et al. Methodologic issues in studies of air pollution and daily counts of deaths or hospital admissions. J Epidemiol Commun Health 50(suppl 1):S3–S11 (1996).
18. Neas LM, Dockery DW, Koutrakis P, Tollerud DJ, Speizer FE. The association of ambient air pollution with twice daily peak expiratory flow rate measurements in children. Am J Epidemiol 141:111–122 (1995).
19. Samet JM, Zeger SL, Berhane K. Particulate Air Pollution and Daily Mortality. Replication and Validation of Selected Studies. The Phase I Report of the Particle Epidemiology Evaluation Project. Cambridge:Health Effects Institute, 1995;123 pp.
20. Delfino RJ, Zeiger RS, Seltzer JM, Street DH. Symptoms in pediatric asthmatics and air pollution: differences in effects by symptom severity, anti-inflamatory medication use and particulate averaging time. Environ Health Perspect 106:751–761 (1998).
21. Peters A, Dockery DW, Heinrich J, Wichmann HE. Medication use modifies the health effects of particulate sulfate air pollution in children with asthma. Environ Health Perspect 105:430–435 (1997).
22. Neas LM, Dockery DW, Burge H, Koutrakis P, Speizer FE. Fungus spores, air pollutants and other determinants of peak expiratory flow. Am J Epidemiol 143:797–807 (1996).
23. Liu LJ, Delfino R, Koutrakis P. Ozone exposure assessment in a Southern California Community. Environ Health Perspect 105:58–65 (1997).
24. Thomas D, Stram D, Dwyer J. Exposure measurement error: influence on exposure-disease relationships and methods of correction. Annu Rev Public Health 14:69–93 (1993).
25. Navidi W, Thomas D, Langholz B, Stram D. Statistical Methods for Epidemiologic Studies of Health Effects of Air Pollution—Research Report No. 86. Cambridge, MA:Health Effects Institute, 1999;57 pp.
26. McMichael AJ, Anderson HR, Brunekreef B, Cohen AJ. Inappropriate use of daily mortality analyses to estimate longer-term mortality effects of air pollution. Int J Epidemiol 27:450–453 (1998).
27. Zeger SL, Dominici F, Samet JM. Harvesting-resistent estimates of air pollution effects on mortality. Ann Epidemiol 10:171–175 (1999).
28. Schwartz J. Lung function and chronic exposure to air pollution: a cross-sectional analysis of NHANES II. Environ Res 50:309–321 (1989).
29. Stern BR, Raizenne ME, Burnett RT, Jones L, Kearney J, Franklin CA. Air pollution and childhood respiratory health: exposure to sulfate and ozone in 10 Canadian rural communities. Environ Res 66:125–142 (1994).
30. Schwartz S. The fallacy of the ecological fallacy: the potential misuse of a concept and the consequences. Am J Public Health 84:819–824 (1994).
31. Peters JM, Avol E, Navidi W, London S, Gauderman WJ, Lurmann F, Linn WS, Margolis H. A study of twelve southern California communities with differing levels and types of air pollution. I: Prevalence of respiratory morbidity. Am J Respir Crit Care Med 159:760–767 (1999).
32. Peters JM, Avol E, Navidi W, London S, Gauderman WJ, Lurmann F, Rappaport E, Gong H, Thomas DC. A study of twelve southern California communities with differing levels and types of air pollution. II: Effects on pulmonary function. Am J Respir Crit Care Med 159:768–775 (1999).
33. Ott W. Concepts of human exposure to air pollution. Environ Int 7:179–196 (1982).
34. Susser M. The logic in ecological. I: The logic of analysis. Am J Public Health 84:825–829 (1994).
35. Mantel N, Haenszel W. Statistical aspects of the analysis of data from retrospective studies of disease. J Natl Cancer Inst 22:719–748 (1959).
36. Miettinen O. Estimability and estimation in case-referent studies. Am J Epidemiol 103:226–235 (1976).
37. Miettinen O. Theoretical Epidemiology: Principles of Occurrence Research in Medicine. New York:John Wiley & Sons, 1985;1–359.
38. Miettinen O. The "case-control" study: valid selection of subjects. J Chron Dis 38:543–548 (1985).
39. Miettinen O. Design options in epidemiologic research. An update. Scand J Work Environ Health 8(suppl 1):7–14 (1982).
40. Wacholder S, McLaughlin J, Silverman D, Mandel J. Selection of controls in case-control studies. I: Principles. Am J Epidemiol 135:1019–1028 (1992).
41. Wacholder S, Silverman D, McLaughlin J, Mandel J. Selection of controls in case-control studies. II: Types of controls. Am J Epidemiol 135:1029–1041 (1992).
42. Wacholder S, Silverman D, McLaughlin J, Mandel J. Selection of controls in case-control studies. III: Design options. Am J Epidemiol 135:1042–1050 (1992).
43. Langholz B, Thomas DC. Nested case-control and case-cohort methods of sampling from a cohort: a critical comparison. Am J Epidemiol 131:169–176 (1990).
44. Langholz B, Goldstein L. Risk set sampling in epidemiologic cohort studies. Stat Sci 11:35–53 (1996).
45. Borgan O, Langholtz B. Risk set sampling designs for proportional hazards models. In: Statistical Analysis of Medical Data: New Developments (Everitt BS, Dunn G, eds). New York:Oxford University Press, 1998.
46. Greenland S, Thomas DC. On the need for the rare disease

assumption in case-control studies. Am J Epidemiol 116:547–553 (1982).

47. Magnus P, Nafstad P, Oie L, Carlsen K, Becher G, Kongerud J, Carlsen KH, Samuelsen SO, Botten G, Bakketeig LS. Exposure to nitrogen dioxide and the occurrence of bronchial obstruction in children below 2 years. Int J Epidemiol 27:995–999 (1998).

48. Cain KC, Breslow NE. Logistic regression analysis and efficient desing for two-stage studies. Am J Epidemiol 128:1198–1206 (1988).

49. Zhao LP, Lipsitz S. Designs and analysis of two-stage studies. Stat Med 11:769–782 (1992).

50. Reilly M. Optimal sampling strategies for two-stage designs. Am J Epidemiol 143:92–100 (1996).

51. Langholz B, Clayton D. Sampling strategies in nested case-control studies. Environ Health Perspect 102(suppl 8):47–51 (1994).

52. Greenland S. A unified approach to the analysis of case-distributional (case-only) studies. Stat Med 18:1–15 (1999).

53. Maclure M. The case-crossover design: a method for studying transient effects on the risk of acute events. Am J Epidemiol 133:144–153 (1991).

54. Khoury MJ, Flanders WD. Nontraditional epidemiologic approaches in the analysis of gene-environmental interaction: case-control studies with no controls. Am J Epidemiol 144:207–213 (1996).

55. Mittleman MA, Maclure M, Robbins JM. Control sampling strategies for case-crossover studies: an assessment of relative efficiency. Am J Epidemiol 142:91–98 (1995).

56. Neas LM, Schwartz J, Dockery D. A case-crossover analysis of air pollution and mortality in Philadelphia. Environ Health Perspect 107:629–631 (1999).

57. Marshall RJ, Jackson RT. Analysis of case-crossover designs. Stat Med 12:2333–2341 (1993).

58. Willich SN, Lewis M, Lowel H, Arntz H-R, Schubert F, Schroder R. Physical exertion as a trigger for acute myocardial infarction. N Engl J Med 329:1684–1690 (1993).

59. Greenland S. Confounding and exposure trends in case-crossover and case-time-control designs. Epidemiology 7(3):231–239 (1996).

60. Bateson TF, Schwartz J. Control for seasonal variation and time trend in case-crossover studies of acute effects of environmental exposures. Epidemiology 10:539–544 (1999).

61. Zaffanella LE, Savitz DA, Greenland S, Ebi KL. The residential case-specular method to study wire codes, magnetic fields, and disease. Epidemiology 9:16–20 (1997).

62. Lee J-T, Schwartz J. Reanalysis of the effects of air pollution on daily mortality in Seoul, Korea: a case-crossover design. Environ Health Perspect 107:633–636 (1999).

63. Greenland S. The effect of misclassification in matched-pair case-control studies. Am J Epidemiol 116:402–406 (1982).

64. Greenland S. Interpretation and choice of effect measures in epidemiologic analysis. Am J Epidemiol 125:761–768 (1987).

65. Breslow N, Day N. The Design and Analysis of Cohort Studies. Statistical Methods in Cancer Research. Vol II. Lyon:International Agency for Research on Cancer, 1987;1–406.